# 中国科学技术大学软件学院

软件工程实验项目环节开题报告

项目名称：留学信息管理与分析系统

## 成 员 名 单 ：罗迪 随哲朝 章文

导 师 ： 张曙 工 程 领 域： 软件工程 领 域 代 码： **085212** 研 究 方 向： 软件系统设计 所 在 院 系： 软件学院

中国科学技术大学软件学院

填表日期： **2018**年 **11** 月 **15**日

### 一、 简况

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 名称 | 中文 | 留学信息管理与分析系统 | | | |
| 英文 | Information Management and Analysis System  for Studying Abroad | | | |
| 项目组成员名单 | 姓名 | | 学号 | 项目中的分工 | 签 章 |
| 罗迪 | | SA18225272 | 信息挖掘与分析 |  |
| 随哲朝 | | SA18225324 | 数据库设计 |  |
| 章文 | | SA18225505 | 系统整体框架设计 |  |
| 中 英 文 摘 要 | 随着留学生人数每年20%的增长速度，留学“大众化”和“低龄化”成为了中国留学生的突出特点，因此如何选择合适的学校和专业成为了许多家长与学生在出国前面临的难题。针对网络上留学网站信息的无序、繁杂，本项目基于互联网的留学信息管理与分析系统应运而生。项目通过抓取国外名校网站与一系列留学服务网站的信息，整合、共享留学系统的信息资源，具体内容包括入学条件、如何进行个人评估、材料准备、申请学校等，使管理分析更加精准，更加便捷地服务出国留学人员。  With the 20% annual growth rate of international students, the “popularization” and “lower age” of studying abroad have become the prominent features of Chinese students. Therefore, how to choose the right schools and majors has become a difficult problem for many parents and students before going abroad. In view of the disorderly and complicated information on the website of the study abroad on the Internet, the Internet-based information management and analysis system for studying abroad came into being. The project integrates and shares the information resources of the study abroad system by grasping the information of foreign famous school websites and a series of study abroad service websites. The specific contents include admission conditions, how to conduct personal assessment, material preparation, and application for schools, so that the management analysis is more precise and more Conveniently serve overseas students. | | | | |
| 主题  词 | 中文 | 数据分析，数据挖掘，web编程 | | | |
| 英文 | data analysis/data mining /web programming | | | |

二、选题依据

**1，**阐述该选题的研究意义，或工程设计的价值和意义，国内外研究现状和发展趋势，选题的先进性和实用性，技术难度及工作量。

近年来，我国出国人数呈逐年上涨趋势，官方数据显示从1978年到2015年，我国派出的各类出国留学人员累总数305.86万人。我国出国留学人员总数不断攀升，虽然增速从2013年开始回调，但是整体人数增长放缓的同时，本科及以下层面就读人数增长却依然迅猛，低龄化趋势愈发明显，学生及其家长对国外院校的要求、对出国留学的诉求也在逐渐发生改变。网上各类留学中介品牌泛滥，缺乏一个稳定、极具公信力的咨询通道；另外，对于院校信息直接的检索和获取得到的内容比较繁杂，留学学生及其家长难以直观、详细地了解国外院校、资格审查等信息。

作为非中介功能系统中的留学生信息管理与分析系统，目的在于结合中国留学现状及前景趋势，对留学现状及趋势进行全面、细致的调研分析，采用定量及定性的科学研究方法给予准备出过留学的学生以科学的个人评估，依托权威数据资源和一手的资料数据是学生完成对于国外院校的具体了解。项目实现的系统能获取各个高校的入学条件，专业排名等信息，学生可以自己选择心仪的高校进行查询，也可以通过科学的个人评估获得一些推荐高校来帮助学生进行选择，系统在在学生申请高校时会显示高校的入学条件是否与学生匹配。该管理系统能够简化学生和家长的选择流程，先进地省去相关人员比较复杂的查询步骤，从而得到一个科学、完善的出国留学方案。

研究管理系统类项目，会涉及到很多管理信息系统（Management Information System，简称MIS）中的知识，即利用计算机硬件、软件、网络通信设备以及其他办公设备，进行信息的收集、传输、加工、储存、更新和维护，以提高效益和效率为目的。实现该系统，志愿留学的学生及其家长会得到一个便捷咨询小助手，直接查询、获取第一线的国外高校资讯。

21世纪以来，自费出国急剧增加，这也导致了一些盲目跟风的行为而国内外的一些留学信息网站信息陈杂，而且有着各种眼花缭乱的广告，并不能为学生提供准确及时的信息，本系统实时爬取国外各大高校的信息，并对其数据进行分析与处理，然后展现给学生，让学生通过信息来判断哪一所学校更适合自己。

留学信息管理与分析系统涉及到许多流行的技术，这些技术会影响到系统判断的准确性，这些关键技术主要有以下几个方面：爬虫技术，利用程序来自动爬取高校的信息，可以通过分布式技术来提高爬虫效率，数据挖掘技术，从大量的数据中通过算法提取出最有价值的内容然后进行分析后展现给用户，web编程技术，设计一个简单易用的界面呈现给用户，让用户能够快速上手，获取自己所需要的信息。

2. 参考文献。

[1] 李江红, 郑海兰. 留学生信息管理系统与办公自动化[C] 。2003.

[2] 白斯古朗. 留学生管理系统分析与设计[D]. 内蒙古大学, 2015.

[3] 乔静. 高校留学生信息管理系统的设计与实现[D]. 南开大学, 2015.布莱恩特，奥哈拉伦. 深入理解计算机系统.机械工业出版社，2011

[4] Bruce Eckel. JAVA编程思想. 机械工业出版社, 2007

[5] 张宗利. 留学生教务管理信息系统的分析与设计[J].2010

[6] 计文柯. Spring技术内幕.第2版[M]. 机械工业出版社, 2012.

[7] Laurie Kutner. Study-Abroad Programs as Information Producers: An Expanding Role for Support of Our Students Studying Abroad[J]. Journal of Library Administration, 2010

三、课题内容及具体方案

**1**．课题内容

本课题主要功能是对留学信息进行爬取，管理与分类，通过数据分析技术对不同人

群进行个性化推荐。

**2．**系统需求分析

国外高校信息的爬取、数据挖掘与分析、数据库设计、系统界面设计。

2.1信息获取

搜索引擎选择谷歌，拟采用python+selenium框架爬取教育部信息网，selenium是一系列基于Web的自动化工具，提供一套测试函数，函数非常灵活，能够完成界面元素定位、窗口跳转、结果比较，能够方便地获取教育部发布的学校名单和学校网址，然后将其信息存入数据库，并根据学校名称和网址获取对应学校的申请条件等信息。

2.2数据挖掘与分析

对数据进行处理，采用Numpy,Pandas库对数据进行清洗与分析，NumPy是高性能科学计算和数据分析的基础包，而 pandas 是基于NumPy 的一种工具，该工具是为了解决数据分析任务而创建的；Pandas 纳入了大量库和一些标准的数据模型，提供了高效地操作大型数据集所需的工具，其中含有大量能使我们快速便捷地处理数据的函数和方法；我们可以将获取到的数据建立一个数据集，通过这两个库可以完成一些数据预处理操作，去除一些脏数据等。

2.3数据库设计

使用mysql数据库，维护四张表，其一是用户表，其二是学校表，其三是学校排名表，其四是专业排名表。

用户表保存用户的登陆信息以及个人信息，诸如留学偏好、成绩信息等。初始登陆信息通过引导用户注册得到，而个人信息则通过让用户在进入系统前进行个人评估获取。

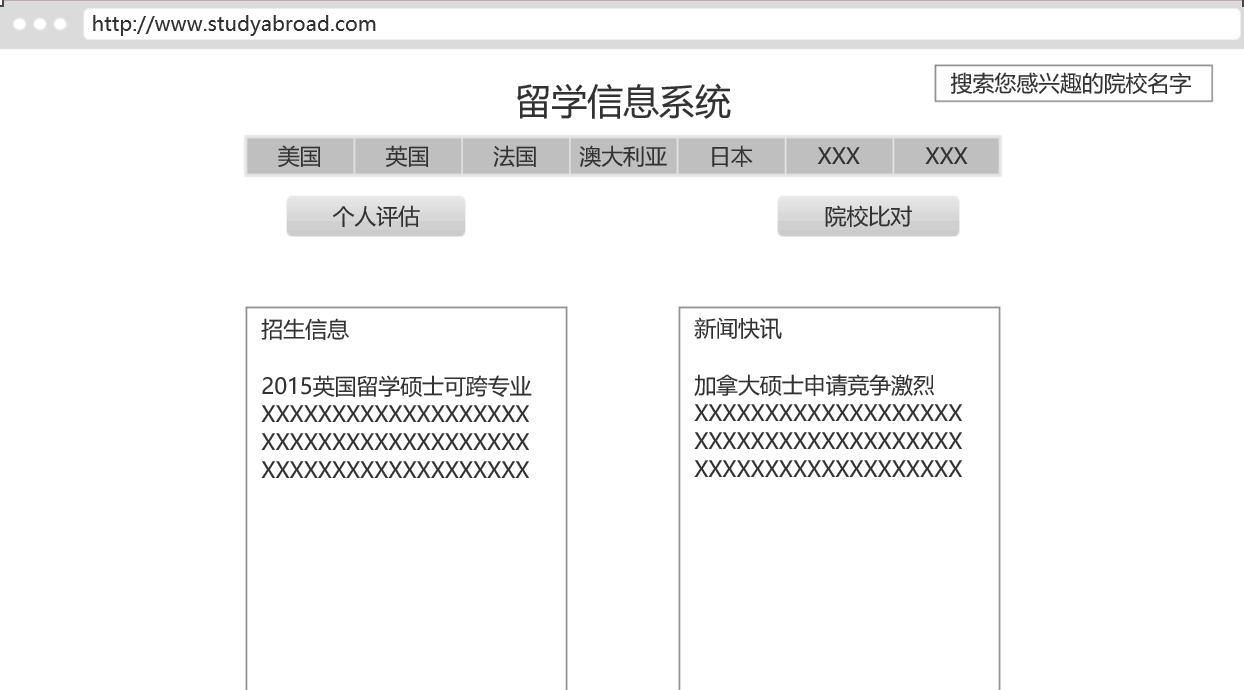
学校表保存学校基本信息、专业信息、申请条件、学费等。学校的信息主要通过爬虫爬取得到。

学校排名表保存学校的综合排名，包括国内排名表和世界排名表。

专业排名表保持各专业的学校排名，为用户提供专业细分方向上的信息。

2.4系统界面设计

系统界面拟采用bootstrap框架设计，系统大概功能界面如下图：



**3**． 系统概要设计

主要包括数据挖掘、数据处理与分析、整体框架设计，系统界面设计

系统在数据库中通过爬虫技术抓取院校信息，用以维护一张院校信息表，信息表中应包括院校名称、所在地、排名情况、申请条件、学费信息等数据项。

用户通过注册用户名密码登入系统，系统提供两个选项，其一：个人评估，用户可以通过填写自身信息得到一份推荐院校名单；其二：理想院校，用户可以在此处填入自己拟定的若干理想院校名录，系统可以生成一份多维度的详细比对表帮助用户选定最终的目标院校。

此外，在界面上方应有按国家标签集聚的院校名录链接，让用户可以快速地根据留学国家获取院校信息。

最后，系统应允许用户通过院校名称搜索该院校的具体信息。

**4**． 拟采用的开发方法、环境，测试方案等

主要利用 JAVA和python在windows 环境下进行开发，并且利用单元测试。

**5**． 技术难度及特色分析

需要掌握web编程及框架的使用，对留学流程有所了解，要会利用爬虫与数据分析技术来获取及分析数据，其中得到的数据格式跟语言都会有很大的不同，如何批量处理这些形态各异的数据是我们需要解决的难题。

四、工作进度的大致安排

预期成果及特色

系统能获取各个高校的入学条件，专业排名等信息，学生可以自己选择所心仪的高校进行查询，也可以通过个人评估获得一些推荐高校来帮助学生进行选择，系统在在学生申请高校时会显示高校的入学条件是否与学生匹配。

整体进度计划分为

里程碑图如下图所示：

里程碑图

具体安排细则：

10/20~11/20 文献调研，进行需求分析，提交开题报告

11/20~11/30 技术路线选择确定，系统细则设计

12/01~12/20 数据爬取、数据分析

12/20~02/01 数据库设计、界面设计

02/01~02/20 版本 1.0 开发完成，形成中期检查报告及相关技术文档

02/20~03/01 添加一些其他功能

03/01~05/01 版本 2.0 的编程以及测试

05/01~05/20 版本 2.0 开发完成

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 导师意见 | 导师签名： | 年 | 月 | 日 |
| 答辩小组意见 | 答辩组长签名： | 年 | 月 | 日 |